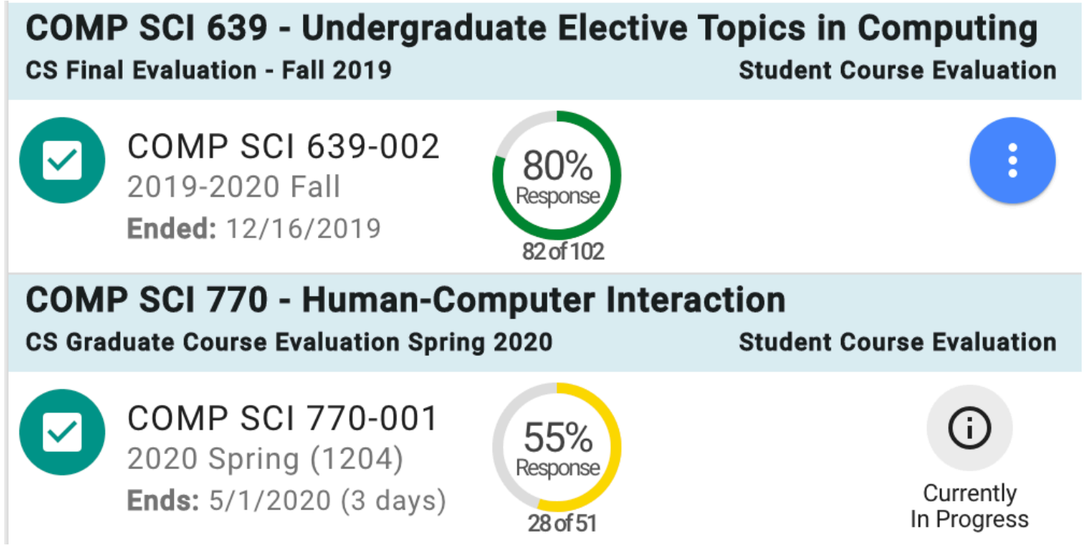
Human-Computer Interaction

Reporting & Writing HCI Papers

Professor Bilge Mutlu

Announcements

- Today is our *last* class
- Please complete the course evaluation (through AEFIS) by **May 1**
  - Response rate is currently at **55%**; last semester was **80%**



Today's Agenda

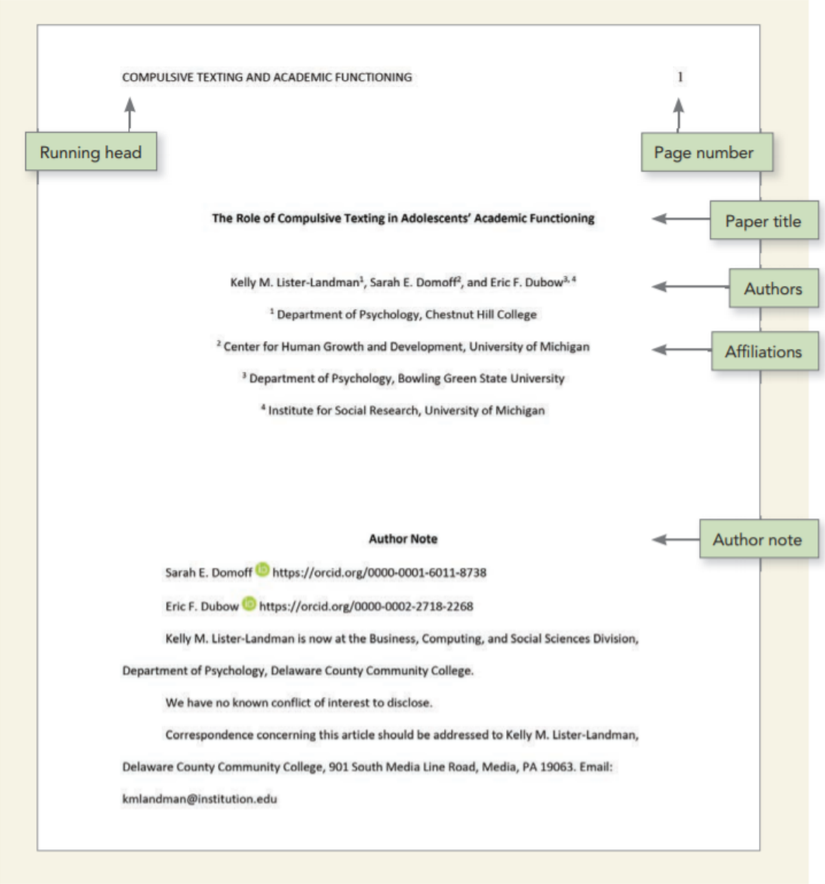
- Overview: *Reporting Statistics, Writing* (30 min)
- Hands-on Activity (20 min)
- Stats session (20 min)

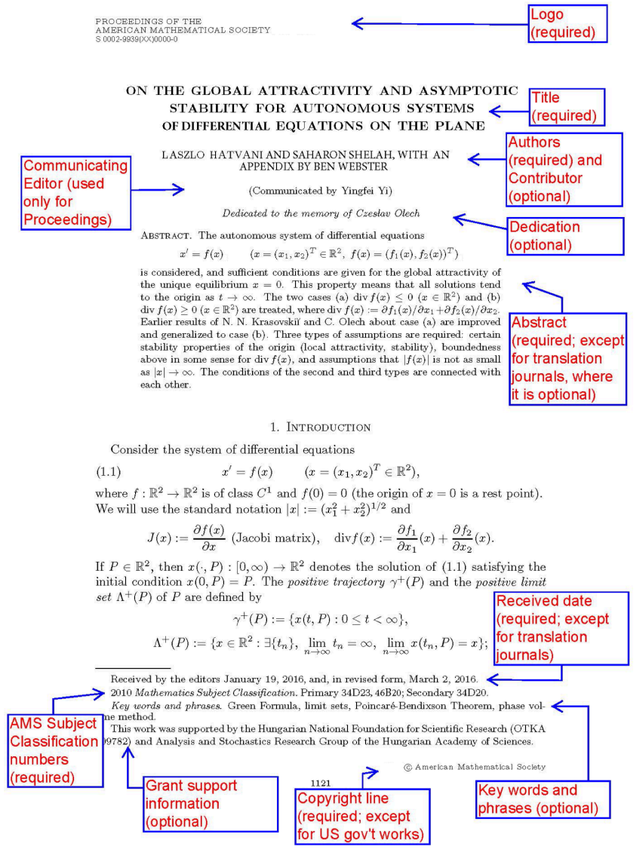
What are reporting norms in HCI research?

Because HCI is a rather eclectic field, the reporting norms are adopted from different fields, roughly as follows:

Aspect	**Norm**
Paper structure	APA (loosely)
Results of statistical analyses	APA (strictly)
Tables, figures	APA (very loosely)
Citations	Depends on the publisher (ACM, IEEE, etc.)
Formulas	AMS (loosely)
Style	APA (loosely), generally high standards in writing

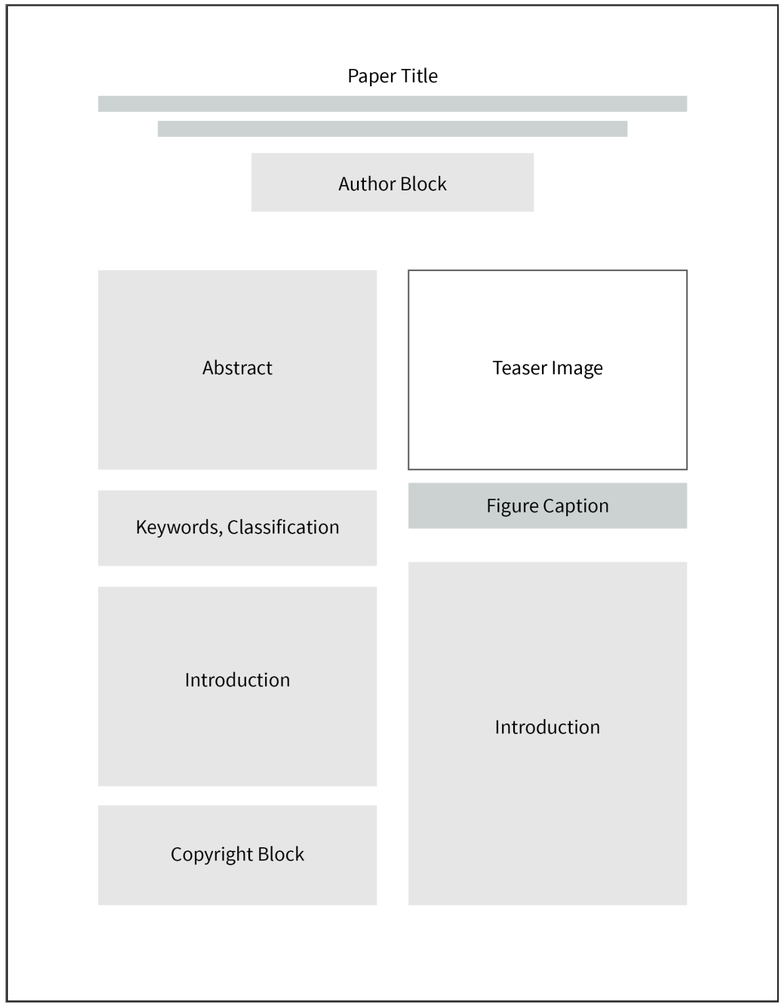
APA Publication Manual: Print, Web; AMS Style Guide: Web[1]

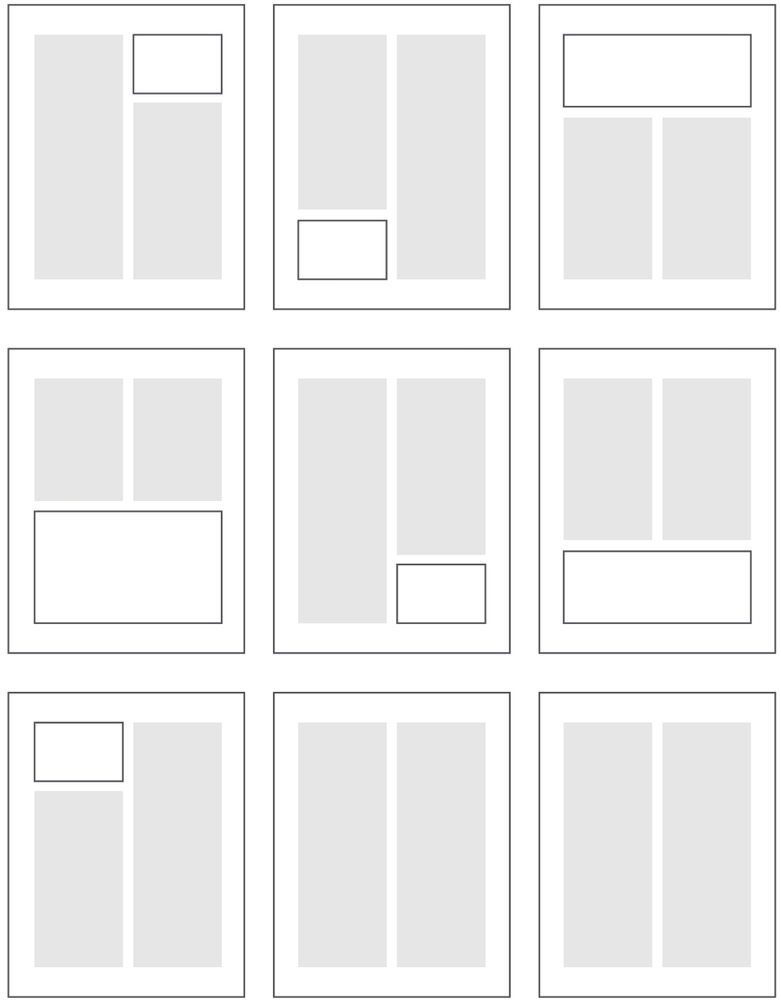




[1] Sources: Left, Right

What does an HCI paper look like?





How is an HCI paper structured?

HCI papers commonly follow the structure below:

- Abstract
- Introduction
- Related Work/Background
- *Hypotheses (quant. empirical)*
- *System/Design (design-based)*
- Method
- Results
- Discussion
- Conclusion
- Acknowledgements
- References
- Appendices

What is an abstract?[2]

The abstract provides a brief but comprehensive summary of the contents of the paper. It gives readers an overview of the paper and helps them decide whether to read the full text. Usually *150 words* max.

The abstract usually includes (1–2 sentences each):

- Summary of literature review
- Problem investigated/RQs
- hypotheses
- methods used
- study results
- implications

[2] APA

How do I choose a title?

There is no formula or requirement, but a few things to consider:

- It should be as short as it can be, but not too broad.
  - E.g., *Bodystorming Human-Robot Interactions*
- A common format in HCI:
  - Catchy headline/System name: Technical title
  - E.g., *Pay attention!: Designing adaptive agents that monitor and improve user engagement*
  - E.g., *Reading socially: Transforming the in-home reading experience with a learning-companion robot*

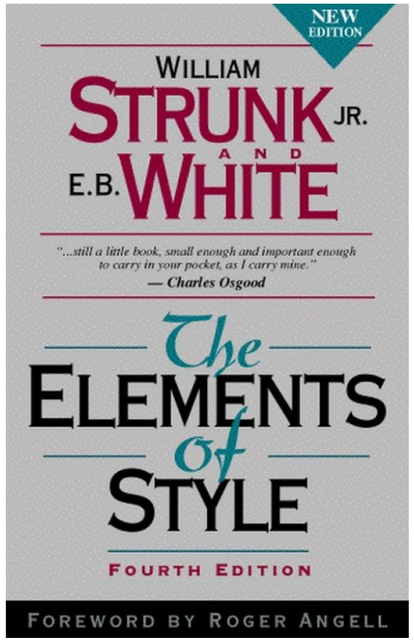
What are other things I should pay attention to?

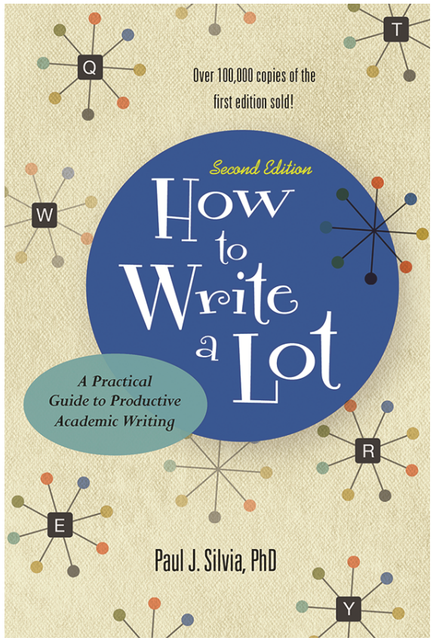
1. Writing
2. Formatting
3. Presentation

Writing[3]

The HCI community pays more attention to writing than most other CS communities, so writing is very important, in particular:

1. Reporting as *storytelling*
2. Flow among parts
3. "Cut deadwood"
4. Avoid any deviation from rules (syntax, grammar, punctuation, etc.)



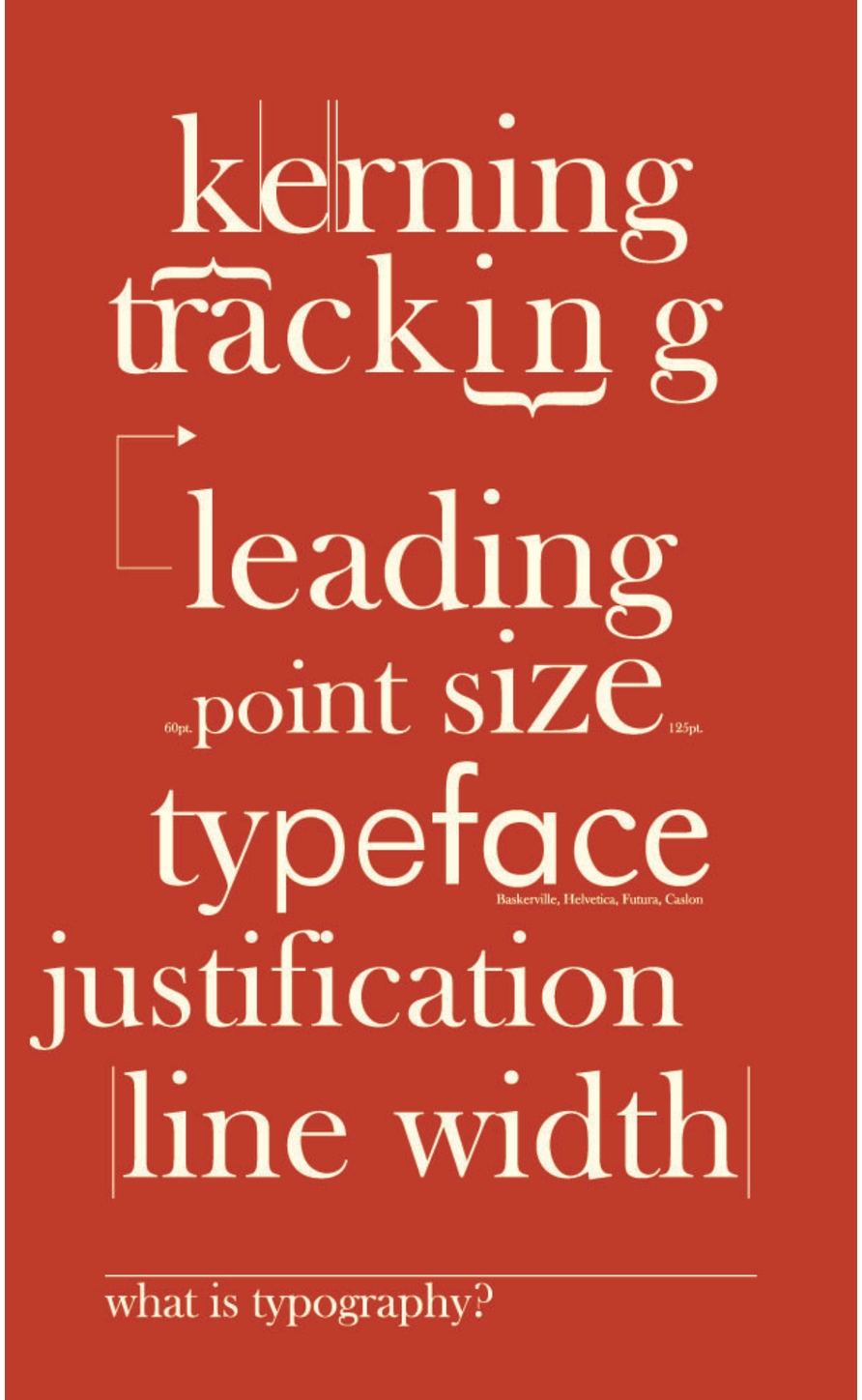


[3] Image sources: Left, Right

Formatting[4]

For good *typography*, become familiar with *leading*, *tracking*, *kerning*, *widows*, *orphans*, *runts*, *rags*, *rivers*.

[4] Image source: Left, Right



Presentation[5]

The overall organization and visual appearance, using informative figures (e.g., a "teaser"), will improve accessibility and appeal.

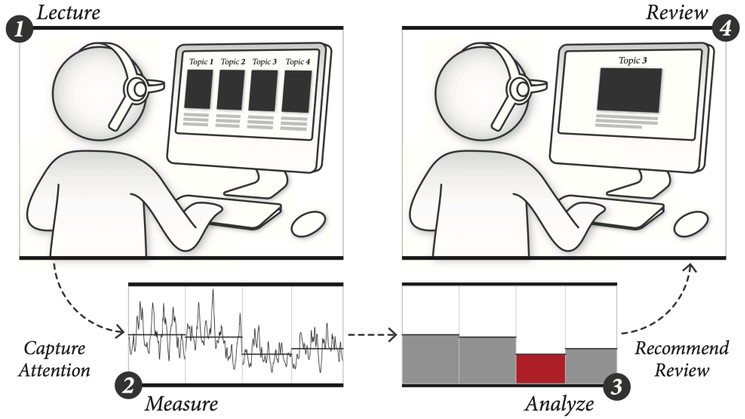


Figure 1. This work presents a novel educational system that (1) instructs users while (2) measuring attention across predefined lesson modules. Following the lesson, the system (3) analyzes the attention measurements to (4) adaptively determine review content that might best improve learning.

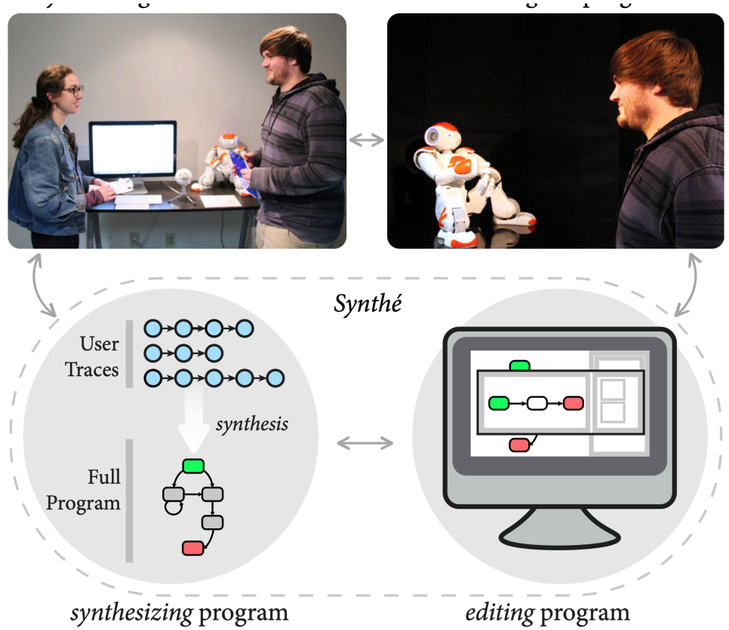
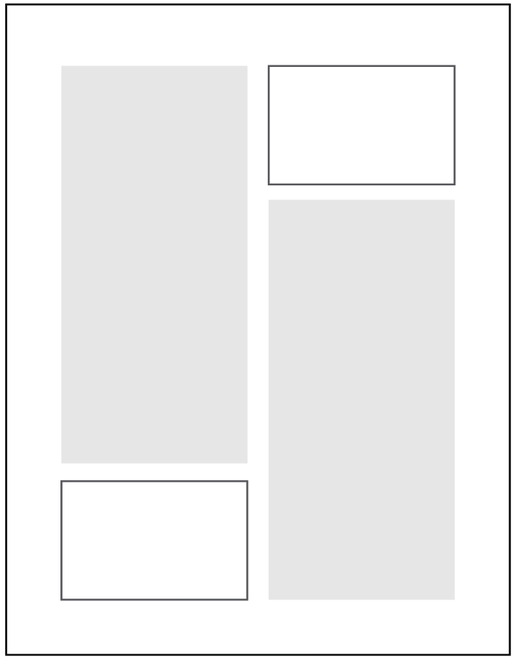


Figure 1. ***Synthé*** **captures designers' demonstrations, synthesizes an in-**



[5] **Left:** Szafir & Mutlu, 2014; **Center:** Porfirio et al., 2019

How do we report statistics?

Descriptive statistics: Distribution characteristics using summary statistics in text, tables, or graphs.

Inferential statistics: Test parameters and results in text or tables and highlighting of significance in graphs.

In *text*, APA guidelines are strictly followed; in *graphs*, you can be creative.

Descriptive statistics[6]

```
> describeBy(data$Guesses, list(data$Leakage,data$TBI))

 Descriptive statistics by group
: Leakage
: HC
   vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
X1    1 291 3.87 1.91      4    3.68 1.48   1  13    12 1.08     1.95 0.11
------------------------------------------------------------------------------
: No Leakage
: HC
   vars   n mean   sd median trimmed  mad min max range skew kurtosis  se
X1    1 367 4.02 1.85      4    3.86 1.48   1  11    10 0.82     0.83 0.1
------------------------------------------------------------------------------
: Leakage
: TBI
   vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
X1    1 282 3.92 2.24      4    3.63 1.48   1  17    16 2.11     7.83 0.13
------------------------------------------------------------------------------
: No Leakage
: TBI
   vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
X1    1 353 4.37 2.46      4    4.05 1.48   1  19    18 1.55     4.24 0.13
```

The healthy controls guessed the item that the robot picked in 3.97 guesses (*SD*=1.91) when the robot gazed toward the item and in 4.02 guesses (*SD*=1.85) when the robot did not gaze toward it. Participants with TBI guessed the robot's pick in 3.92 guesses (*SD*=2.24) when the robot gazed toward it and in 4.37 guesses (*SD*=2.46) when the robot did not.

[6] Data from Mutlu et al., 2018, Social-cue perception

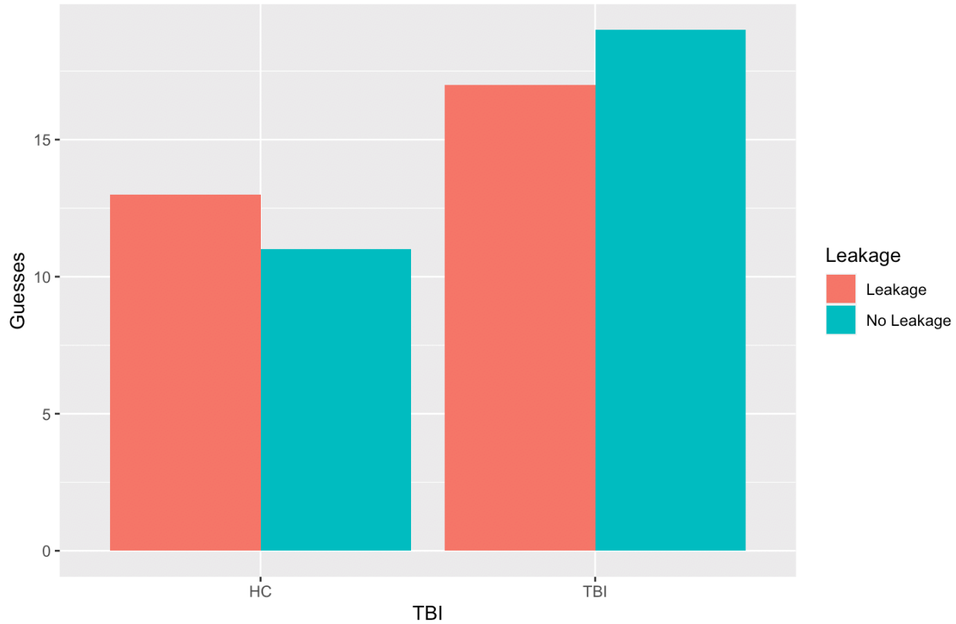
How do we deal with decimals?[7]

For numbers...	Round to...	SPSS	Report
Greater than 100	Whole number	1034.963	1035
10 - 100	1 decimal place	11.4378	11.4
0.10 - 10	2 decimal places	4.3682	4.37
0.001 - 0.10	3 decimal places	0.0352	0.035
Less than 0.001	As many digits as needed for non-zero	0.00038	0.0004

[7] Source

Descriptive statistics (visual)[8]

```
library(ggplot2)
ggplot(data, aes(fill=Leakage, y=Guesses, x=TBI)) +
    geom_bar(position="dodge", stat="identity")
```



[8] More information on using `ggplot2`

Inferential statistics[9]

```
> summary(aov(Guesses~(TBI*Leakage)+Error(ID/Leakage)+TBI,data=data))

Error: ID
             Df Sum Sq Mean Sq F value Pr(>F)
TBI           1   15.2  15.236   2.360  0.127
Leakage       1    4.0   4.012   0.621  0.432
TBI:Leakage   1    7.5   7.467   1.157  0.284
Residuals   142  916.6   6.455

Error: ID:Leakage
             Df Sum Sq Mean Sq F value Pr(>F)
Leakage       1   27.3  27.268   6.680 0.0107 *
TBI:Leakage   1    7.1   7.131   1.747 0.1884
Residuals   144  587.8   4.082
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Error: Within
            Df Sum Sq Mean Sq F value Pr(>F)
Residuals 1001   4325   4.321
```

A mixed-model analysis of variance (ANOVA) revealed a significant effect of the leakage cue, $F(1,144) = 6.68$, $p = .011$.

Participants correctly identified the robot's pick on an average of 3.89 questions ($SD = 2.08$) when the robot displayed the gaze cue and 4.19 ($SD = 2.17$) when it did not.

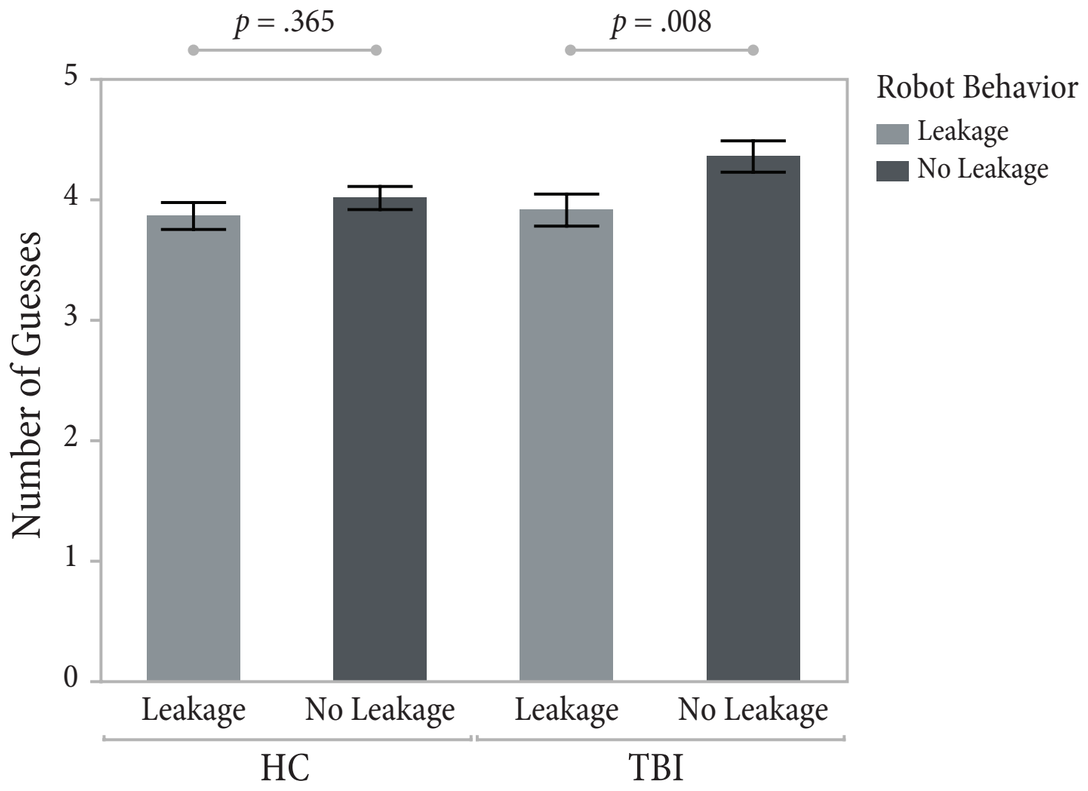
[9] Shown is a simplified model using data from Mutlu et al., 2018

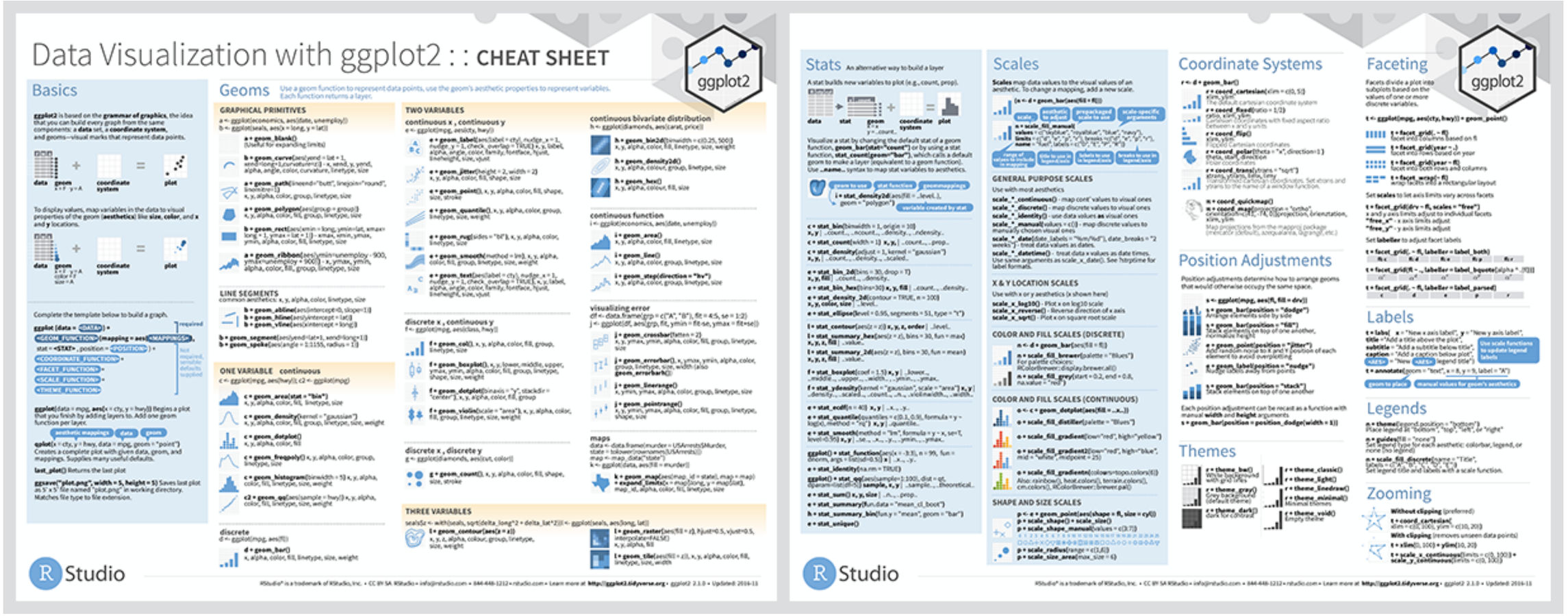
How do I report different tests?[7]

Statistic	Example
Mean and standard deviation	$M = 3.45$, $SD = 1.21$
Mann-Whitney	$U = 67.5$, $p = .034$, $r = .38$
Wilcoxon signed-ranks	$Z = 4.21$, $p < .001$
Sign test	$Z = 3.47$, $p = .001$
t-test	$t(19) = 2.45$, $p = .031$, $d = 0.54$
ANOVA	$F(2, 1279) = 6.15$, $p = .002$, $\eta_p^2 = 0.010$
Pearson's correlation	$r(1282) = .13$, $p < .001$

[7] Source

Test results can also be mapped on graphs either manually (e.g., using Adobe Illustrator) or automatically using advanced scripting (e.g., `ggplot2`, `matplotlib`).





[10] ggplot2 cheat sheet